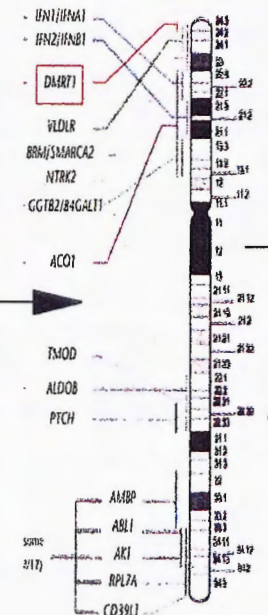
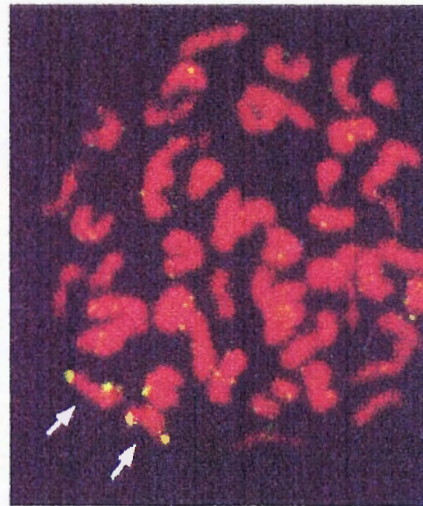


# Génomique fonctionnelle des Tilapias

## J.F. Baroiller et H. D'Cotta

ccatccatcat acgacttact atagggtctg agcggccgoc cgggcagggtg cagttgtctg ccatccagc  
 tctccacccg gcatcttcat ccatctctgt ttatcaatc ttgacacat gtccaaggga ccagctgttg  
 gttatgatc cgggaccaco tactctgttg taggtgtctt caacatgga aaagtigaaa ttatcgtaa  
 tgaccaagga aacaggacila cccaagttaigtgctt acagatactg agnagttag ttgagatgcc  
 gggaaaaatc aggtggccat gaacccccc aacacagctc ttgatgccaa cgtctgaat ggcaggcagt  
 tggatgatgg cgtgttcag ttgatata agcactggcc ttthaatc atcaatgaca attccgtcc  
 caaggtccag gttgaatac aggtggaatc caagtcttc taccotgaag aaattctc tatgtctc  
 accaagatga aggaatgtc tggagctac ctgggaaga cgtgtccaa cgtgtgac accgtctc  
 cgtactca cgaclcaaa agacaggcca ccaaggatg tggtaacatc totgcttga atgtctctg  
 gatcaaat gaaccaact ctgtctcat tcttatgt ttggcaaaa aggttgggc ttgagaaat  
 gtctgatt ttgatcttg tgggtctct ttgatgtt ccatctac catgaggat ggcactctg  
 aggtcaaat tactctgga gacactact tgggagaga agacttgc aacogtatg tgaaccatt  
 catcacagag tcaaacgca agcacaagaa agacatcag gacacaaga gactgttc tctctccg  
 acagctgtg agaggccaa gctaccctg tctcaagca ctacggcag ttatgaatc gactccct  
 atgaggcat cgaclctat acttccatca ccaggcccg cttgaggag ctcaatgct acotctccg  
 tggacattg gaccagtg agaggcact tctgagcc aagatgaca aggtcagat ccatgacatt  
 gctgtgtg gttgtccac tgcattcc aaatccaga agctgtcca agactctc aatggcaagg  
 agctcaaca gacttaac ccagatgagg ctgtgtc ttgagatgct gtcaggctg ccatctctc  
 tggagcaag ttgagaaag ttgagactt gotttgtt gatgtact ctctgtcc ttgaattgag  
 accgtgtcg gactatgac tttctcat aagcttaaa caacatcc aaccaaac actcagact  
 tcaactta ctggacaac cagctgtgtg tctcatca ggtctatg ggtgaacgt caatgacaa  
 ggacaacaa ttgtggga aattgagt cactggaat ccactgcac ctgtgtgt cctcaaat  
 gaggctact ttgactga ttccatgg atcatgaat ttacgtgt ttgacaagag actggcaagg  
 agaacaat caclatcat aatgataag gctgtcag caaggagac attgagcga ttgtcagga  
 agctgagaag tacaaggctg aggatgat gtagctgac aaggtttg ccaagaacg tctgagct  
 tatgtctta acatgaagto cactgtgga gatgagaag tgaaggcga gatcagtgac gaggacaac  
 agangattc ttgacattgt aatgaagica tggctgtgt ttgacaagac cagactgca agagagaaga  
 gttgagcac cagcaagagg agctggaga ggtatgaac ccatcatca ctgaagctga ccagatgtct  
 gaggatgct ctgtgggaat gcccgaagg atgcccgtg gattccagg tctgtgtct gctcaggctg  
 gttgatctc cggccaac atcaggagg tgaactaag cgtatccga aagccactat cgtactca  
 tacaatgt tactgtgct ctctgtgt gaactctt aatgtatt actactga aaggtttg  
 aggactcaa ctgagaga gatgttga attataaaa gaaagggga ttaaggatc acacttca  
 atcacaggga acagacttc tactagatt gcacgtta actgtgac gttcaag ctttttt ttttttt  
 tccaactgc ttgatctt gaaggacaag aataaagg ttacttacc cc



**Gestion de Biodiversité**  
**Amélioration génétique**  
**Sélection sur génotype**



# **Génomique:** Acquisition systématique de données de structure et de fonction sur l'ensemble du génome

## **Pour quoi faire ?**

Identifier des gènes impliqués dans des caractères d'intérêt zootechnique

Mieux gérer les ressources génétiques

Accroître l'efficacité des programmes d'amélioration génétique

## **Comment ?**

**Génomique structurale:**

Balisage  
génomique par des  
marqueurs  
locus-spécifiques

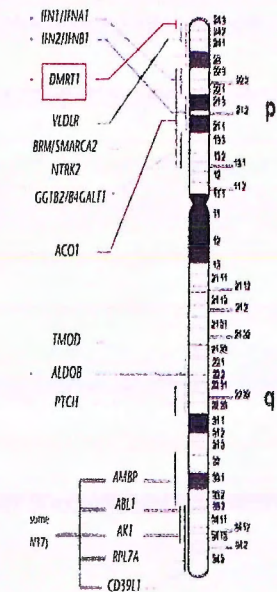
Séquençage  
systématique

**Génomique expressionnelle:**

gènes exprimés  
dans un tissu

**Génomique fonctionnelle:**

traduire séquences  
en données  
biologiques



# La Génomique, pour quoi faire ?

La génomique consiste en l'acquisition systématique de données de structure et de fonction sur l'ensemble du génome.

La finalité d'une telle approche est triple :

- ❑ Identifier des gènes impliqués dans des caractères d'intérêt zootechnique,
- ❑ Mieux gérer les ressources génétiques,
- ❑ Accroître l'efficacité des programmes d'amélioration génétique

Trois niveaux sont généralement considérés :

1) La génomique structurale :

C'est le balisage du génome par des marqueurs locus-spécifiques et séquençage systématique

2) La génomique expressionnelle :

C'est la recherche des gènes exprimés à un temps t, dans un tissu donné, et dans une condition physiologique précise

3) La génomique fonctionnelle :

elle cherche à traduire des séquences en données biologiques



## La Génomique en France

"Ceux qui n'auront pas su ou pas pu dans les prochaines années s'assurer la propriété ou la copropriété des outils et du matériel vivant nécessaires aux améliorations animales et végétales se trouveront en situation de dépendance"

- Message de Claude Allègre, rencontre des Directeurs d'Unité INRA, 22 Avril 1999, Paris-Bercy





## Politique Nationale et Génomique

Le gouvernement a décidé d'apporter un soutien particulier aux recherches sur le génome, qui recevront plus de 2 milliards de francs sur 3 ans. Interview de C. Allegre , Le Monde, 22/09/99.

## Politique Scientifique du CIRAD et Génomique

« L'environnement dans lequel le CIRAD évolue connaît des changements considérables : la mondialisation, la génomique, .... Nous avons su prendre les grands tournants, comme ...., la participation à Génoplante.... Quant à la génomique, ..., le CIRAD ne peut en être absent, car il risquerait de priver ses partenaires d'éventuelles retombées positives ». Interview de B. Bachelier, CIRAD-Infos, Octobre 1999.

Un petit nombre de laboratoires publics et privés est en train de se partager la connaissance du génome de quelques organismes vivants, et en particulier pour les productions animales, le tilapia. Les biotechnologies associées à une meilleure connaissance du génome, « englobent des techniques de portée suffisamment générale pour trouver des applications utiles à toutes les catégories de producteurs, qu'ils bénéficient ou non de compétences techniques, de capitaux, d'un environnement économique performant. C'est le rôle des organismes publics, et du CIRAD en ce qui le concerne, de veiller à ce que des applications soient aussi développées pour d'autres bénéficiaires potentiels<sup>1</sup>. » C'est aussi le rôle du CIRAD, d'informer objectivement (puis éventuellement de former) les partenaires du Sud, des avantages et inconvénients (risques environnementaux, géopolitiques, économiques et sociaux liés aux biotechnologies associées).

---

<sup>1</sup> CIRAD-Infos, juillet-Aout 1999



## Situation de La Génomique des Poissons

Organismes	Classification	N chrom	Taille du génome n (Mb)
<i>Homo sapiens</i>	Mammifères	23	3500
<i>Mus musculus</i>	Mammifères	20	3000
<i>Salmonidés</i> <i>Oncorhynchus mykiss</i>	Poissons	26-48* 29-33	" 2400
<i>Tilapias: Oreochromis niloticus</i>	Poissons	22	" 1200
<i>Drosophila melanogaster</i>	Insectes	4	160
<i>Caenorhabditis elegans</i>	Nématodes	6	100
<i>Arabidopsis thaliana</i>	Plantes	5	100
<i>Saccharomyces cerevisiae</i>	Levures	16	14
<i>Escherichia coli</i>	Bactéries	1	4.8



## **Quelques caractéristiques sur les génomes des poissons d'intérêt aquacole**

Si l'on excepte le projet en cours sur le génome humain, un nombre réduit d'espèces animales ou végétales, fait ou fera l'objet de tels programmes. Parmi les espèces de Poissons d'intérêt aquacoles, la famille des Salmonidés et le groupe des tilapias ont été retenus à la fois compte tenu des enjeux économiques, ou sociaux économiques (2 des 3 groupes d'espèces majeures pour l'aquaculture), scientifiques (voir plus loin, les études sur la spéciation et l'évolution des espèces réalisées dans les grands lacs africains sur les Cichlidés, famille des Tilapias) mais aussi des outils spécifiques disponibles dans les laboratoires internationaux.



## Projets et Compétences Génomique Tilapia + Truite

Titre / Espèces	Labo.	Avancement
Genetic Maps of Aquaculture species Tilapia, Truite, Poisson-chat	Thomas Kocher Univ. New Hampshire, USA	?
Tilapia Genome Project	Thomas Kocher Univ. New Hampshire, USA	162 marqueurs
Genome Mapping, QTL Tilapia (salinité, résistance au froid)	Univ. Davis, USA + Agricultural Research Organization, Israel	112 marqueurs
Cartographie génome Tilapia	Biosoft (société privée)	300 marqueurs (500 attendus)
Truite	Univ. Guelph, USA	220 marqueurs (objectif 300)
Truite arc-en-ciel, T. commune,	INRA	
Truite	Univ. Washington	400 marqueurs
saumon pink	Univ. Montana	620



## Contexte International et Génomique

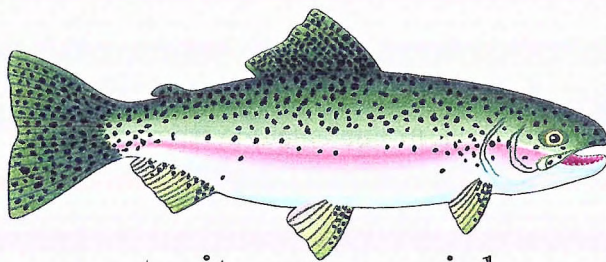
Pour les Productions aquatiques, la Connaissance du Génome et ses applications associées (recherche de QTL, sélection sur génotype, ...) font d'ores et déjà partie des priorités des programmes de travail de la plupart des bailleurs de fonds internationaux et/ou européens. La majorité de ces programmes concerne essentiellement les salmonidés et les Tilapias.

Concernant le tilapia, des programmes de génomique structurale (essentiellement cartographie) sont en cours dans des structures publiques aux USA (Univ. du New Hampshire et à l'Univ. C. Davis aux USA), au Canada (Univ. de Dalhousie), en Israël (Agricultural Research Organization), au Canada, et même en Europe (Univ. de Stirling en Ecosse), ou privées (Biosoft en Norvège). Une bonne partie du génome du tilapia est donc déjà bien balisé par des marqueurs moléculaires souvent disponibles dans des banques de données consultables sur le réseau et/ou commercialisés par des structures privées comme Biosoft (Norvège).

# Projet européens débutants ou déposés en 1999

## ■ Projet **ASTEROGER: INRA**

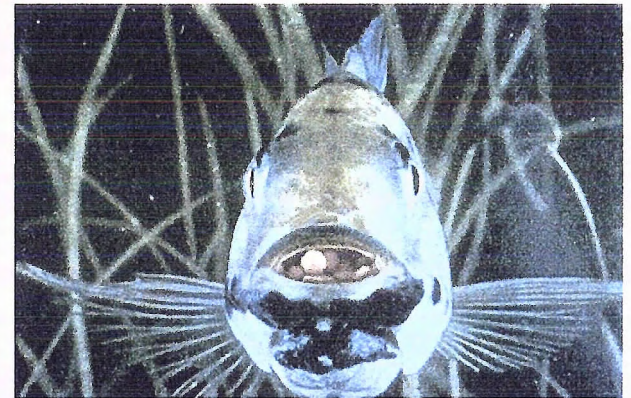
- ▶ bovins
- ▶ porcins
- ▶ poulet
- ▶ truite



truite arc-en-ciel

## ■ Projet **FARM GENE** déposé à la CEE

- ▶ bovins
- ▶ porcins
- ▶ poulet
- ▶ tilapia (Biosoft = partenaire privé Norvégien, Univ. Stirling (Ecosse), Univ. Liège (Belgique), Univ. de Nijmegen (Hollande))





# Les Projets Européens ou Français de Génomique intégrant le modèle poisson

## Europe (en cours)

Programme européen de cartographie comparée sur les salmonidés (SALMAP), participation de l'INRA-Génétique des Poissons à Jouy-en-Josas

## France (Démarrage en Automne 1999)

ASTEROGER, « Analyse Systématique des Transcrits en vue de l'Etablissement d'un Répertoire Ordonné de Gènes pour l'Etude leur Régulation » est un grand projet transversal associant les départements de Physiologie Animale, Santé Animale, Elevage et Nutrition, Génétique, et Hydrobiologie et Faune Sauvage de l'INRA.

Quatre espèces animales d'élevage sont prises en compte :  
un bovin, le porc, le poulet, et la **truite**.

La DG de l'INRA a attribué 1,5 MF pour le démarrage du projet intégrant les 4 espèces ; la première débute dès l'automne 1999, et devra être achevée en juillet 2000.

## Europe (soumis pour financement à la Commission Européenne)

FARM-GENE, « Towards a functional map of the farm animal genome »

Quatre espèces animales d'élevage sont prises en compte :  
un bovin, le porc, le poulet, et le **tilapia**.

Labos impliqués pour le tilapia : Univ. de Stirling en Ecosse, Université de Liège en Belgique, Univ. de Nijmegen en Hollande, et un partenaire privé Biosoft en Norvège

## **Génome fonctionnel TRUITE (INRA)**

- 1) Construction d'une banque ADNc multi-tissus  
Normalisée et organisée (96 ou 384 puits)  
  
Obtention de 50.000 ESTs**
- 2) Séquençage systématique de 25.000 ESTs  
(CNS ~ 40 MF= 4 x 50.000 séq.) et gestion**
- 3) Construction de filtres haute densité, microarray  
ou puce (Pôle technique à Rennes ??)**
- 4) Localisation d'ESTs sur le génome avec panels  
d'hybrides d'irradiation (*Génétique Poissons, Jouy-en-Josas*;  
300 marqueurs)**
- 5) Fonction des gènes potentiels et localisation sur le génome**



## Génomique Fonctionnelle chez la truite (INRA)

Cette étude s'inscrit dans le cadre d'un programme plus large, ASTEROG<sup>\*</sup>, intéressant l'ensemble des productions animales.

“ Analyse Systématique des Transcrits en vue de l'Etablissement d'un Répertoire Ordonné de Gènes pour l'Etude de leur Régulation ”

*(Voir programme détaillé en annexe)*

L'INRA commence dès l'automne 1999, le projet ASTEROG<sup>\*</sup> pour la caractérisation des génomes de 4 espèces d'animaux d'intérêt agronomique:

- bovin
- porc
- poulet
- **truite**

Le but est l'obtention pour chaque espèce d'un répertoire des gènes transcrits sous la forme d'EST comprenant et représentant une grande partie du génome codant de cette espèce.

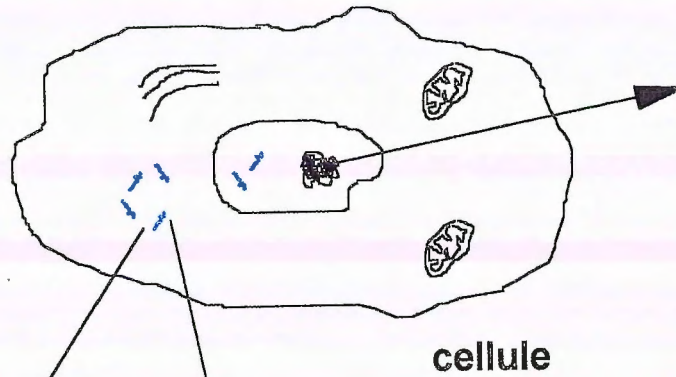
Pour cela, cinq étapes majeures sont mises en œuvre à partir de l'automne 1999 : Construction pour chaque espèce, d'une banque multi-organes d'ADNc, normalisée et ordonnée, séquençage, construction de filtres haute densité, localisation des EST, recherche de la fonction des gènes correspondants.

L'intérêt et le principe général de ces différentes techniques et étapes sont rappelés dans les illustrations suivantes.

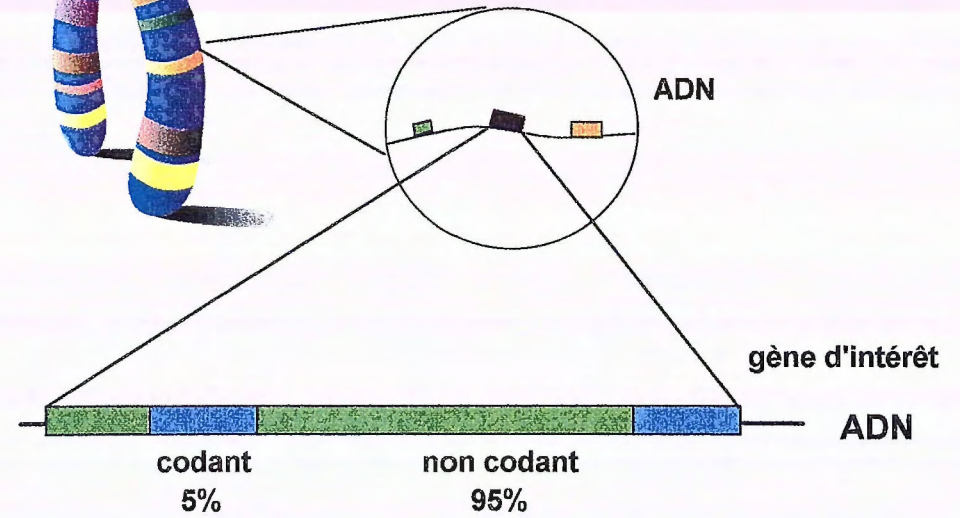
# GENOME

10 à 20.000 gènes  
exprimés par  
cellule

ARNm  
(cellule spécifique)



chromosome

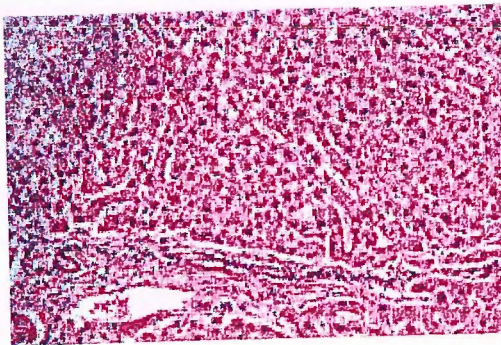


Génome: 100.000 gènes

MUSCLE

GONADE

FOIE





# Génome Expressionnel ou Fonctionnel

L'étude du Génome Expressionnel ou Fonctionnel cible les parties codantes du génome, riches en information.

Le génome d'un mammifère possède environ 100.000 gènes mais seulement 10 à 20.000 de ces gènes sont exprimés, sous forme d'ARNm, dans chaque type cellulaire.

Ces ARNm ou transcrits sont spécifiques:

- ☐ à chaque cellule
- ☐ à un stade de développement
- ☐ à une condition physiologique donnée.

# GENOME

gène d'intérêt



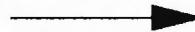
- 5% de séquences codantes
- travailler sur séq. codantes
- Régions riches en information
- Gènes exprimés dans un tissu au moment du prélèvement

**ARNm**

**ARNm**  
(cellule spécifique)



RT



ADNc



*primer S*

*primer AS*



**Amplification d'ADNc  
PCR**





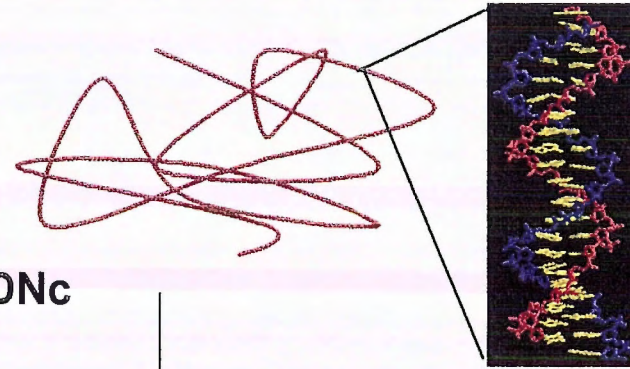
## **La transcription inverse = passage du transcrit à l'ADN correspondant**

Il est possible de remonter à l'ADN complémentaire (ADNc) et donc à la copie codante des gènes à partir des ARNm, par une réaction de transcription inverse.

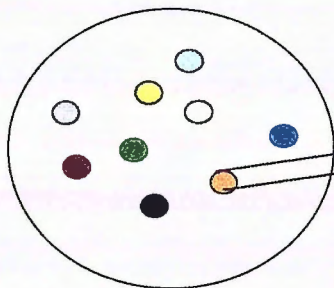
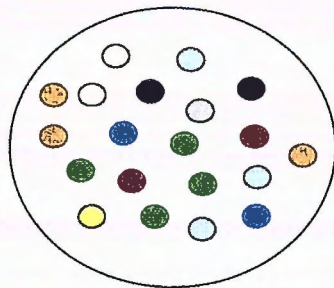
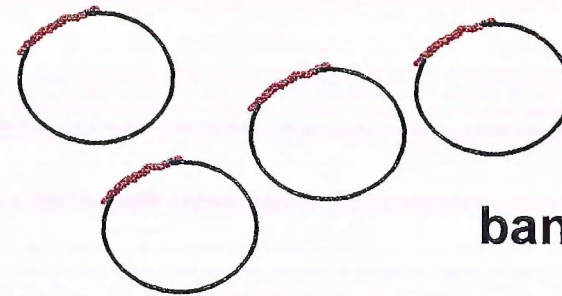
Ces ADNc peuvent être ensuite amplifiés de manière exponentielle par la technique classique de la PCR (polymerase chain reaction).

# GENOME TRUITE (INRA)

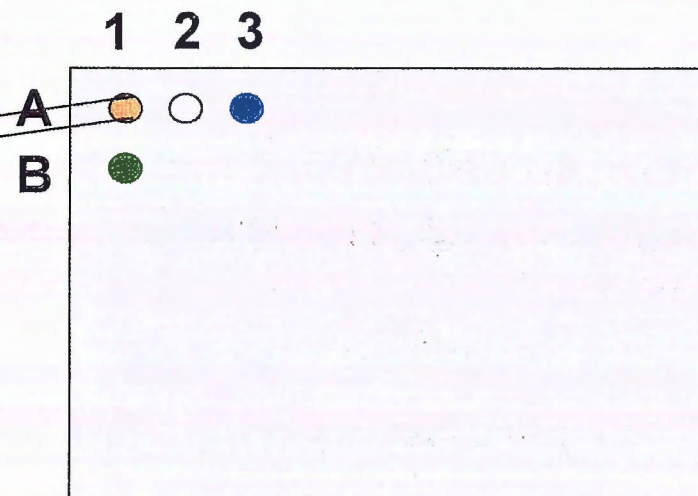
ADNc



banque d'ADNc



banque normalisée



banque ordonnée



## **ASTEROGER : les premières étapes**

(Analyse Systématique des Transcrits en vue de l'Etablissement d'un Répertoire Ordonné de Gènes pour l'Etude de leur Régulation : *Voir programme détaillé en annexe*)

Les premières étapes de ce projet sont:

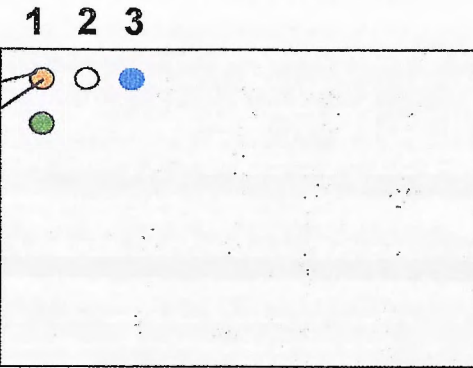
- ❑ Extraction des ARNm des différents tissus pour la construction d'une banque d'ADNc multi-tissus
- ❑ Cette banque d'ADNc sera normalisée (élimination de toute redondance pour ne conserver qu'une seule copie par transcrit).
- ❑ La banque sera ensuite ordonnée dans des plaques de 96 ou 384 puits, permettant une localisation facile (et un retour possible) et rapide de chaque transcrit de la banque.

# EST (Etiquettes)

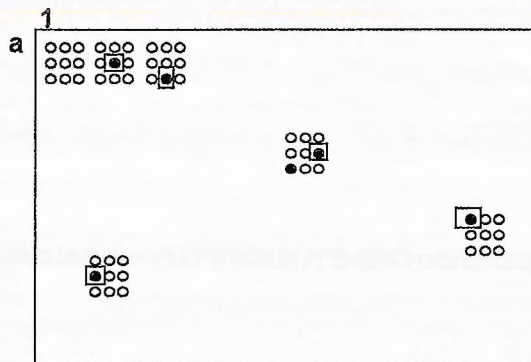
gène hsp90

EST- 5'

EST- 3'



banque ordonnée



Filtres à haute densité  
ou microarray

sonde  
(tissu/condition A)

Séquençage (CNS)

Base de données



## **L'INRA envisage de créer 50.000 EST par espèce.**

Les EST (expressed sequence tags) ou étiquettes, ont été utilisées dans le cadre du projet du génome humain

- Ce sont des fragments des extrémités 5' et 3' d'un gène, de 100 à 500 paires de bases.
- C'est une méthode extrêmement rapide et peu coûteuse permettant le séquençage systématique des gènes sans la nécessité du séquençage complet de chacun d'entre eux.  
Dans un premier temps, L'INRA séquencera 25.000 EST par espèce grâce à des accords avec le CNS (Centre National de Séquençage) ou/et d'autres structures de séquençage.
- Les EST permettent la recherche des homologues de séquence dans les banques de données et éventuellement d'attribuer une fonction à un gène.

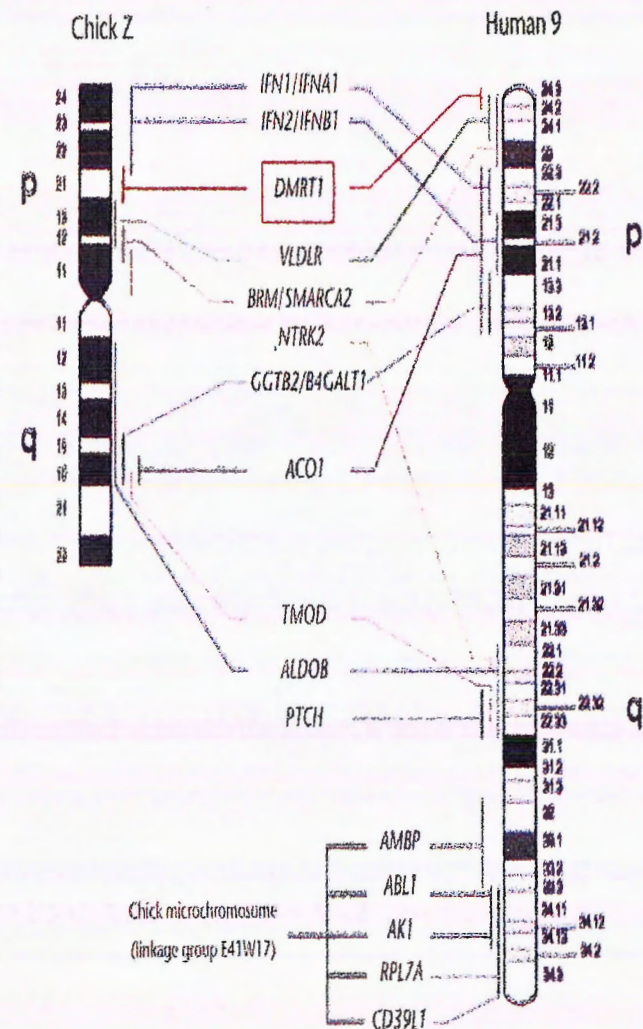
L'analyse simultanée de l'expression de nombreux gènes en utilisant des EST constitue l'une des approches majeures de la génomique expressionnelle.

- Pour cela, ces EST sont organisées et "micro-déposées" avec impression robotique sur des supports divers: membrane de nylon (utilisables à plusieurs reprises), verre ou silicone (augmentation de la densité). Ces filtres ou microarray peuvent contenir des densités conventionnelles ou des hautes densités avec des milliers de clones répartis sur une très faible surface.
- Les supports seront hybridés, pour comparer les profils d'expression des EST, en utilisant des sondes complexes obtenues à partir des ARNm des différents tissus ou dans conditions physiologiques d'intérêt. Une ou plusieurs sondes d'ADN peuvent être marquées par de la radioactivité ou avec des fluorochromes.
- L'acquisition des données après hybridation est faite par des systèmes robotiques: densitomètre, analyse d'image, logiciels ....

# Conservation de Synténie entre le chromosome Z de Poulet et le chromosome 9 humain (300 millions d'années de divergence)

- Nanda, Shan, Schartl et al., 1999 (Nature genetics)

**Fig. 1** Comparative location of orthologous genes on chicken Z and human chromosome 9. The G-banded idiogram of chicken (GGA) Z chromosome (left) opposite an idiogram of G-banded human (HSA) chromosome 9 (right). Comparatively mapped genes are indicated between the idiograms. The distal long arm of HSA 9q32-qter appears to be homologous to chicken linkage group E41W17 on a microchromosome. Human mapping information was obtained from OMIM.





## **Transférabilité des acquis de la génomique d'une espèce à une autre**

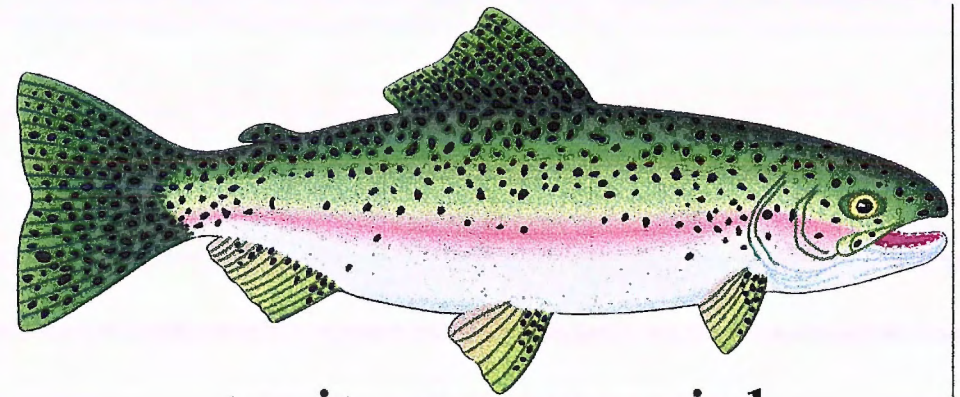
Approche comparative: toutes les études faites sur le génome de truite devront idéalement être transférables à d'autres espèces, en particulier au tilapia.

Récemment Nanda *et al.*, 1999 ont démontré une conservation de groupes de synténie (groupes de liaison entre 2 marqueurs moléculaires) du chromosome Z de poulet et du chromosome 9 humain.

Considérant que la divergence évolutive entre ces deux espèces est d'environ 300 millions d'années, l'hypothèse d'une certaine homologie truite/tilapia entre au moins certains groupes de liaison (voire entre certaines EST) reste raisonnable.

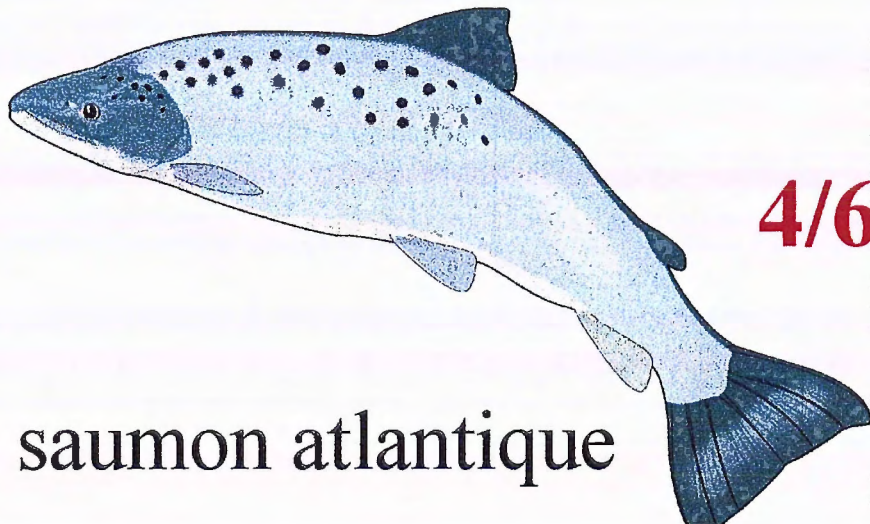
# Conservation des **groupes de Liaison** entre différentes espèces de Salmonidés

- Gharbi, Guyomard et al., 1999;
- Gautier, Gharbi, Krieg, Guyomard et al., 1999



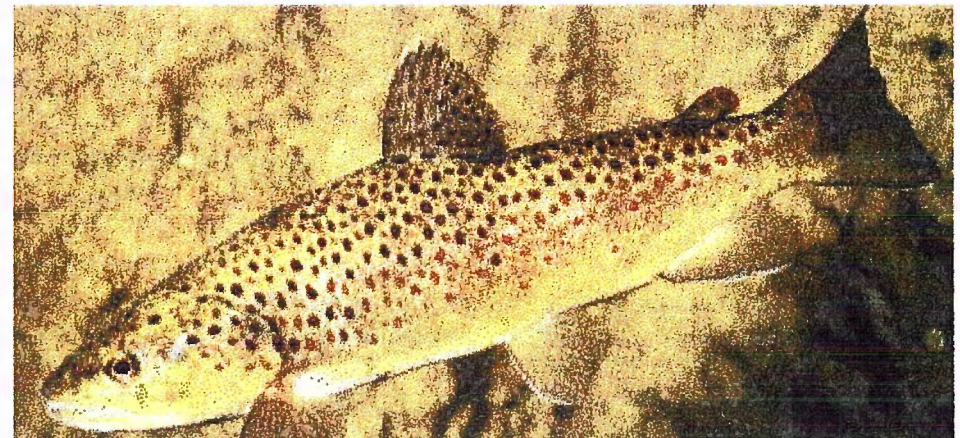
truite arc-en-ciel

**9/13**



saumon atlantique

**4/6**



truite commune



## **Exemple de transférabilité entre espèces de poissons d'une même famille**

Le groupe de génétique des poissons (INRA, Jouy-en-Josas) a mis en évidence la conservation d'une grande proportion de groupes de liaison entre la truite arc-en-ciel et la truite commune (9 groupes conservés sur 13), ainsi qu'entre la truite commune et le saumon atlantique (4 groupes conservés sur 6).

Là encore, l'hypothèse d'une certaine homologie truite/tilapia entre au moins certains groupes de liaison (voire entre certaines EST) reste raisonnable.

## Génome fonctionnel TILAPIA (CIRAD)

- **Tilapia**: Gène d'intérêt différent (comportement parental, facteurs mineurs du déterminisme du sexe, gènes thermosensibles, ...), ou plus facilement identifiable que ceux de la truite (pas de tetraploidisation)
- Caractères sous le contrôle d'autres gènes: gène d'importance différent
- Transférabilité de truite à tilapia: ~ 50% information dûe à l'homologie des parties des séquences
- Souches et espèces à caractéristiques extrêmes (Physio-pathologie)
- Générations courtes (bon modèle pour la sélection)
- Gestion de la Biodiversité des tilapias et des Cichlidés dont Génotypage faits sur *O. niloticus* et *S. melanotheron*, et leurs hybrides

Familles de référence



## **Exemple de transférabilité entre espèces de poissons d'une même famille**

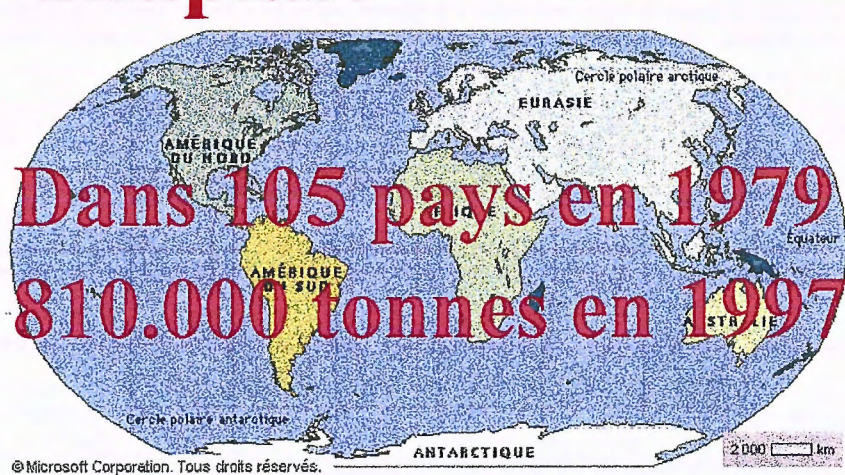
Le groupe de génétique des poissons (INRA, Jouy-en-Josas) a mis en évidence la conservation d'une grande proportion de groupes de liaison entre la truite arc-en-ciel et la truite commune (9 groupes conservés sur 13), ainsi qu'entre la truite commune et le saumon atlantique (4 groupes conservés sur 6).

Là encore, l'hypothèse d'une certaine homologie truite/tilapia entre au moins certains groupes de liaison (voire entre certaines EST) reste raisonnable.

# Le Tilapia, modèle complémentaire pour la Génomique Poissons

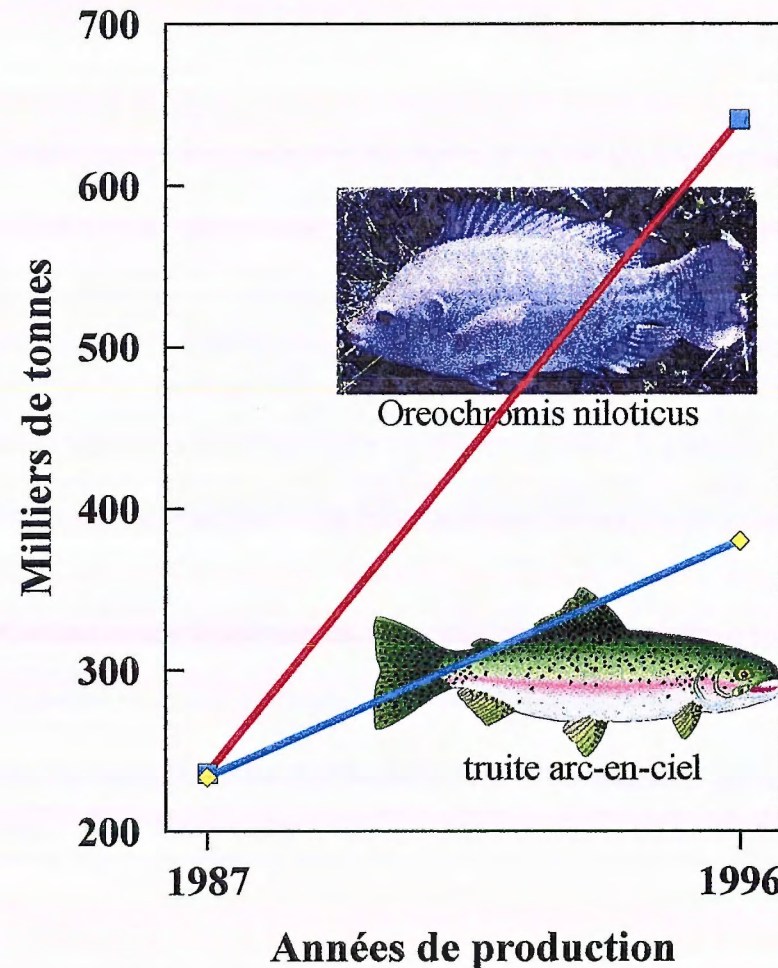
## ■ Productions mondiales de tilapia, *O. niloticus*

### ■ Tilapias:



- ▶ **2ème** groupe d'espèces de poissons produits en **Eau douce**
- ▶ **3ème** groupe d'espèces aquacoles (**E. douces + marines**)

## Evolution de la production aquacole d'*O. niloticus* et de la TAC





# **Le Tilapia Modèle complémentaire pour la génomique Poissons**

## **Le tilapia, une espèce majeure pour l'aquaculture et le développement**

Tous les Programmes de génomique reposent sur des réseaux de compétences et des pôles techniques. Le Projet Génomique Tilapia doit s'intégrer :

- à l'intérieur de dispositifs en place (Laboratoire de Génétique des Poissons de l'INRA pour la cartographie du génome) ou démarrant (Station SCRIBE à Rennes, pour la composante poisson du programme de génomique fonctionnelle, ASTEROGGER, démarrant dès l'automne 1999).
- Et/ou à l'intérieur de pôles techniques et de compétences (Génoplate, et futur Génopôle à Montpellier).

Deux types d'arguments sont pris en compte, par le CIRAD (comme par l'INRA, partenaire sur cette espèce sous réserve d'une implication CIRAD) pour affirmer l'intérêt d'un modèle complémentaire pour des approches de génomique :

1. les enjeux économiques, socio-économiques et politiques
2. les apports/originalités scientifiques

### **1. Enjeux socio-économiques :**

Les tilapias (810.000 tonnes produites en 1997) constituent le 2<sup>ème</sup> groupe d'espèces de poissons produits en eau douce, et le 3<sup>ème</sup> groupe toutes eaux confondues (eaux douces et marines). L'importance de cette production des tilapias s'explique par leurs nombreux avantages zootechniques dont une extrême tolérance à des milieux et conditions d'élevage difficiles, leur faible coût de production et leur vaste répartition géographique qui en découle (ce groupe est présent dans plus de 105 pays suite à de nombreuses introductions dans le passé). En Afrique, comme en Asie, c'est l'une des 2 espèces d'eaux douces les plus consommées.



# Le Tilapia, modèle complémentaire pour la Génomique Poissons

## ■ Interactions Comportement parental / Cycle sexuel

INRA, ORSTOM, Univ. Rennes-CNRS, Univ. Liège

### Enjeu appliqué + Intérêts du modèle:

#### Tilapias

- avec Incubation buccale
  - maternelle stricte
  - paternelle stricte
  - biparentale
- sans Incubation

#### Inhibition cycle sexuel / Comportement parental

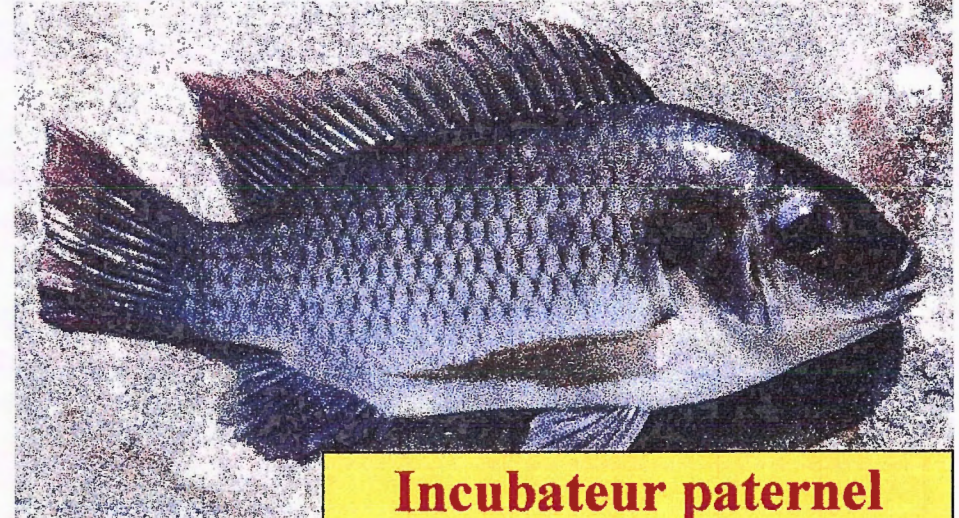
#### Modèle / Spéciation (Cichlidés des Gds Lacs Africains)

*Oreochromis niloticus*



**Incubateur maternel**

*Sarotherodon melanotheron*



**Incubateur paternel**



## **Le Tilapia Modèle complémentaire pour la génomique Poissons**

### **2. Le tilapia et ses spécificités biologiques : Existence d'un comportement parental parfois très élaboré**

Les tilapias, comme beaucoup de Cichlidés présentent toute une palette de comportements parentaux développés, allant d'une simple garde des œufs ventilés dans leurs nids par un courant d'eau généré par des mouvements de nageoires, à une incubation buccale des œufs puis des alevins.

Selon les espèces, l'incubation buccale est pratiquée exclusivement par le père, par les 2 parents ou exclusivement par la mère. Dans ce dernier cas, l'incubation buccale inhibe le cycle sexuel suivant, qui ne pourra débiter qu'après l'arrêt de la garde des alevins par leur mère.

La comparaison d'espèces de tilapia où l'ovogenèse n'est pas influencée par un tel comportement (espèce à incubation paternelle) et d'espèce ou au contraire, l'incubation buccale inhibe la reprise de la vitellogenèse, constitue un modèle de choix pour mieux comprendre et contrôler la régulation de l'ovogenèse. Or pour la plupart des espèces d'aquaculture, il convient de contrôler le cycle sexuel, soit pour induire des reproductions ou les synchroniser, soit au contraire pour empêcher toute reproduction anarchique.

Il en est de même quant à l'étude des relations entre la complexité/efficacité du comportement parental (et donc de la survie larvaire) et la réduction du nombre d'œufs pondus.

*Une thèse co-encadrée par l'INRA et le CIRAD a été réalisée sur ce domaine.*

# Le Tilapia, modèle complémentaire pour la Génomique Poisson

## ■ Dimorphisme de Croissance lié au Sexe: **ATP-CIRAD**

**Enjeu appliqué + Intérêts du modèle:**

Génotypes sexuels (XX, XY, YY, ...)

Croissance rapide = exp. courtes

Sensibilité facteurs externes, f. sociaux



**Collaborations**

**France**

**INRA, ORSTOM**

**Europe**

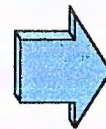
**Pays de Galles, Belgique**

**Afrique**

**Côte d'Ivoire, Niger, Burkina-Faso**

**Asie**

**Philippines**



**Pérennisation sujet + réseau**

**Outils moléculaires spécifiques en cours d'acquisition**

**Montage Projet CEE RD INCO-DC**



# **Le Tilapia Modèle complémentaire pour la génomique Poissons**

## **2. Le tilapia et ses spécificités biologiques : Existence d'un Dimorphisme de croissance lié au sexe**

Chez le tilapia, comme chez beaucoup d'autres espèces d'intérêt aquacole (bar, turbot, perche, ...), il existe un dimorphisme de croissance lié au sexe ; il convient alors soit de privilégier le sexe présentant la meilleure croissance (contrôle du sexe), soit au contraire de rechercher les conditions dans lesquelles, le sexe opposé peut néanmoins présenter des croissances acceptables.

L'existence d'espèces à forte croissance présentant un dimorphisme important, et inversement d'espèces à faible croissance où le dimorphisme n'existe pas ou peu, et la possibilité de réaliser des hybrides viables et fertiles entre ces 2 espèces, constitue également un modèle original, pour l'analyse du déterminisme de cette croissance différentielle.

Par ailleurs, de nombreux outils biologiques spécifiques (populations génétiquement mâles ou femelles permettant d'obtenir dès la fécondation des populations de sexe-ratios connus, possibilité de disposer de tous les génotypes sexuels possibles,...) permet d'étudier l'influence des facteurs sociaux et génétiques sur les caractéristiques de ce dimorphisme.

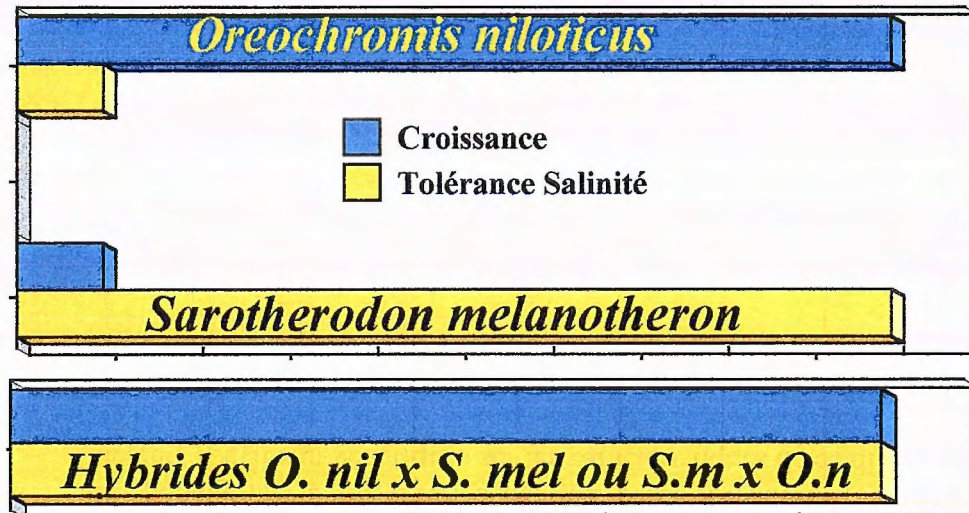
*L'ensemble de ces caractéristiques a permis en effet d'analyser et de mieux contrôler le dimorphisme de croissance lié au sexe, par l'intermédiaire d'une ATP-CIRAD, dans laquelle l'INRA intervenait comme l'un des partenaires majeurs.*



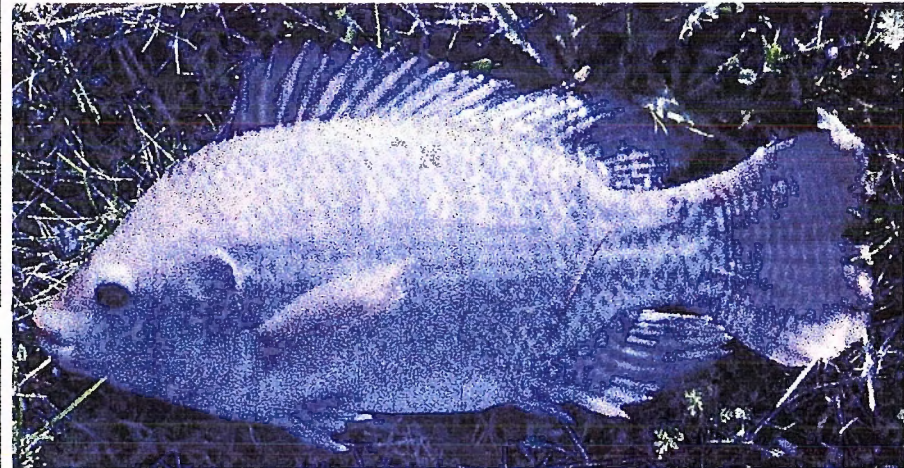
# Le Tilapia, modèle complémentaire pour la Génomique Poissons

## ■ Tolérance à la salinité: ATP-HIT Hybrides Intergénériques de Tilapia

INRA, IFREMER, IRD, Univ. Liège, Philippines, Bénin



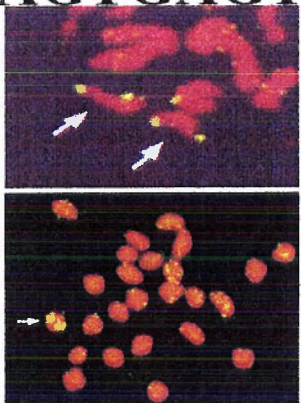
*Oreochromis niloticus*



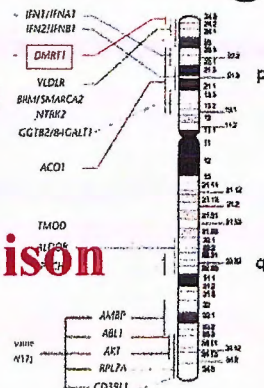
*Sarotherodon melanotheron*



ctgtgtgtgAGTGAGTGAGTGAGTG  
AGTGAGTGAGTGAGTGacagttgg



**µsatellites**  
**FISH**  
**Grpes Liaison**





# Le Tilapia Modèle complémentaire pour la génomique Poissons

## 2. Le tilapia et ses spécificités biologiques : la résistance à la salinité

La vaste aire de répartition de ce groupe d'espèces reflète entre autres une importante capacité d'adaptation à de nombreuses conditions et milieux difficiles. L'existence d'espèces capables de s'adapter à de très fortes salinités (110‰, soit plus de 3 fois la salinité de l'eau de mer), d'espèces moins tolérantes (20‰) et la possibilité de produire des hybrides viables et fertiles entre ces espèces, permet aussi bien d'envisager des programmes de sélection génétique pour associer en une même souche, croissance et résistance à la salinité, que d'étudier le déterminisme de cette adaptabilité à de tels milieux.

*L'ensemble de ces caractéristiques a permis de mettre en place un programme d'hybridation inter-générique et de sélection génétique visant à la création d'une telle souche à forte croissance et forte résistance à la salinité, pour répondre à une demande de diversification des milieux d'élevage du tilapia. Au sein de cette ATP-CIRAD HIT dans laquelle intervient l'INRA, des marqueurs moléculaires (microsatellites) sont développés et/ou optimisés chez le tilapia. Ces marqueurs sont ceux majoritairement utilisés en génomique structurale pour baliser le génome et établir des cartes.*

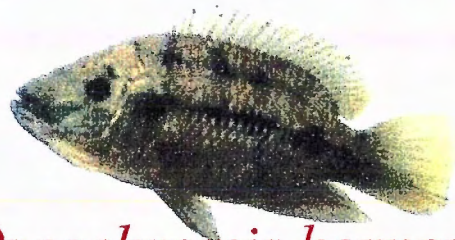
# Le Tilapia, modèle complémentaire pour la Génomique Poissons

## ■ Déterminisme Génétique du Sexe

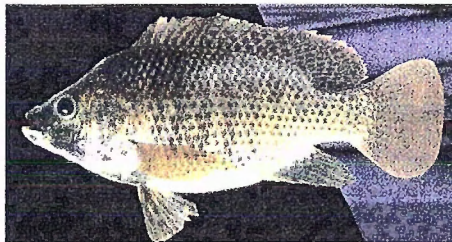
Projet CEE FAIR, + collaboration INRA Génétique (B. Chevassus, D. Chourrout)

### Modèle Oiseaux

Femelle ZW / Mâle ZZ



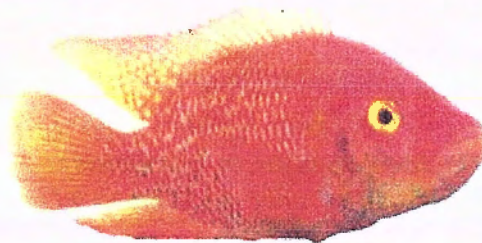
*Oreochromis hornorum*



*Oreochromis aureus*

Hybrides fertiles

XXZ, XW, YZ, YW



Génotypes en  
intraspécifique ds 2 sexes

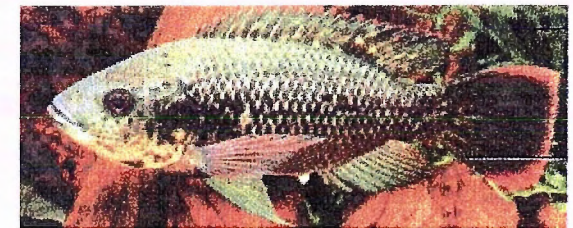
XX, XY, ZZ, ZW,  
YY, WW

### Modèle Mammifères:

Femelle XX / Mâle XY



*Oreochromis niloticus*



*O. mossambicus*



# Le Tilapia Modèle complémentaire pour la génomique Poissons

## 2. Le tilapia et ses spécificités biologiques : co-existence de déterminisme du sexe de type Mammifères et de type Oiseaux au sein de ce groupe d'espèces

Dans le groupe des tilapias, 2 types de déterminisme génétique du sexe sont identifiés :

Une partie des espèces de tilapia possèdent un déterminisme du type de celui des mammifères, le mâle étant XY et la femelle XX, Les autres espèces présentent au contraire un déterminisme du sexe du même type que celui des Oiseaux, les mâles étant ZZ et les femelles WZ.

Compte tenu de la possibilité d'hybrider ces espèces entre elles, il est possible de disposer de tous les génotypes sexuels de type Mammifères et de type Oiseaux ainsi que de tous leurs intermédiaires.

Enfin, chez les Poissons, comme chez tous les Vertébrés inférieurs, il est possible d'inverser le sexe gonadique et phénotypique d'un individu par des traitements hormonaux ou thermiques pendant la période indifférenciée. Dès lors, des individus, viables et fertiles possédant des génotypes soit opposés à leur phénotype (mâles XX, femelles XY, mâles ZW, femelles ZZ) soit nouveaux (n'existant théoriquement pas dans le milieu naturel) peuvent être produits (mâles YY, femelles WW).

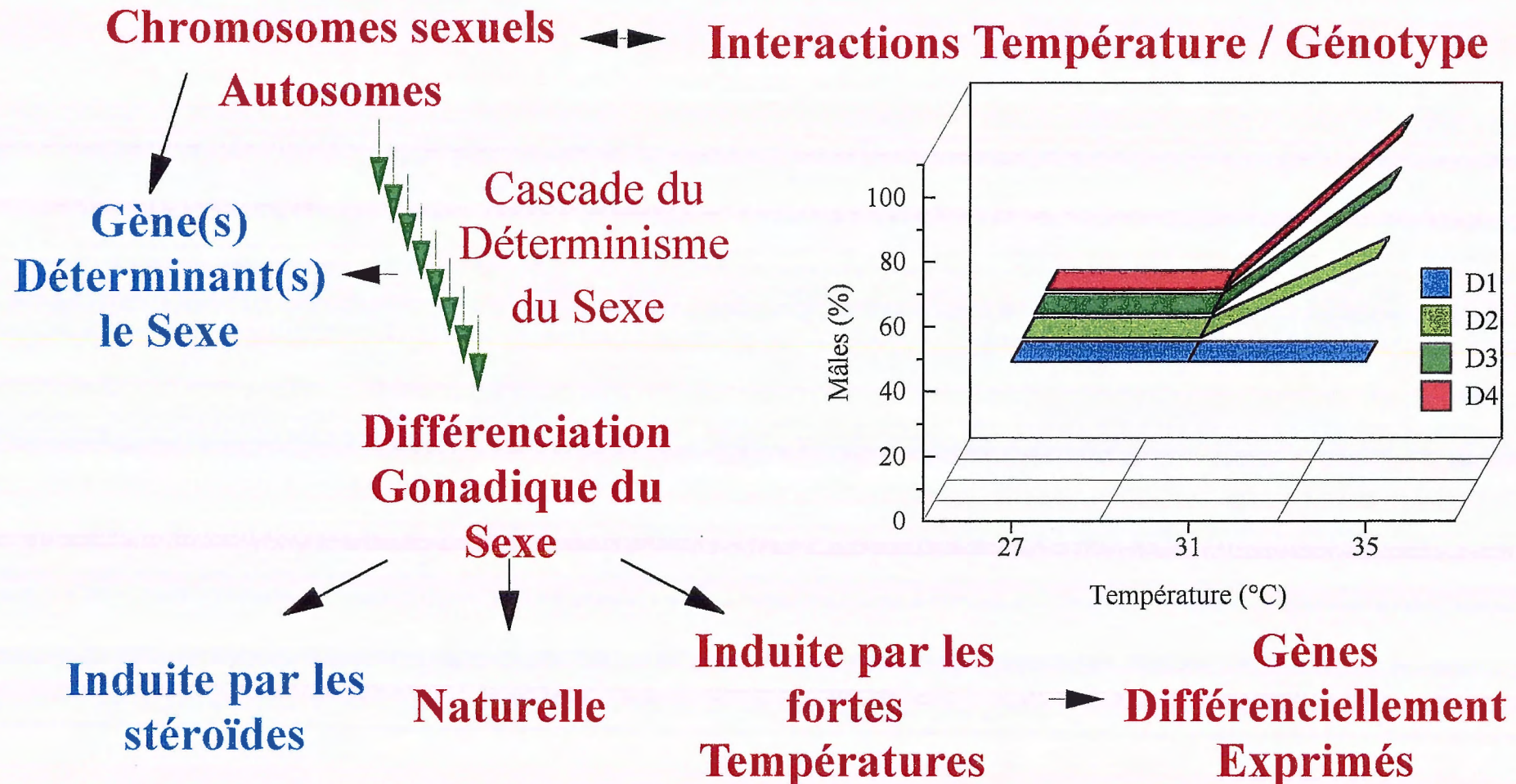
Ces caractéristiques font du tilapia, un modèle idéal pour étudier le déterminisme du sexe et son évolution dans le règne animal. En outre, l'utilisation de tels génotypes permet la production de populations monosexes mâles (ou monosexes femelles) recherchées par les pisciculteurs pour leur forte croissance (cf le dimorphisme de croissance lié au sexe, et la grande efficacité de reproduction du tilapia).

*L'ensemble de ces caractéristiques sont à la base d'un des volets d'un **projet européen FAIR**, coordonné par le CIRAD et dans lequel l'INRA est l'un des partenaires avec son modèle truite.*

# Le Tilapia, modèle complémentaire pour la Génomique Poissons

## ■ Thermosensibilité de la Différenciation du Sexe des Tilapias

Projet CEE FAIR: 6 partenaires européens - Coordination CIRAD-EMVT (JFB)





# **Le Tilapia Modèle complémentaire pour la génomique Poissons**

## **2. Le tilapia et ses spécificités biologiques : sensibilité de la différenciation gonadique du sexe aux températures élevées**

Chez le tilapia, les fortes températures peuvent masculiniser une partie au moins de certaines descendances. La thermosensibilité de ces descendances dépend de facteurs parentaux, paternels comme maternels. Ainsi le sexe-ratio de certaines descendances n'est pas affecté par de fortes températures, inversement des populations 100% mâles sont obtenues par une simple élévation de quelques degrés de la température pendant une courte période critique.

Les effets des facteurs de l'environnement sur la différenciation du sexe ont été jusqu'ici peu étudiés, et constituent une approche nouvelle et originale pour la compréhension des mécanismes de régulation de la différenciation du sexe.

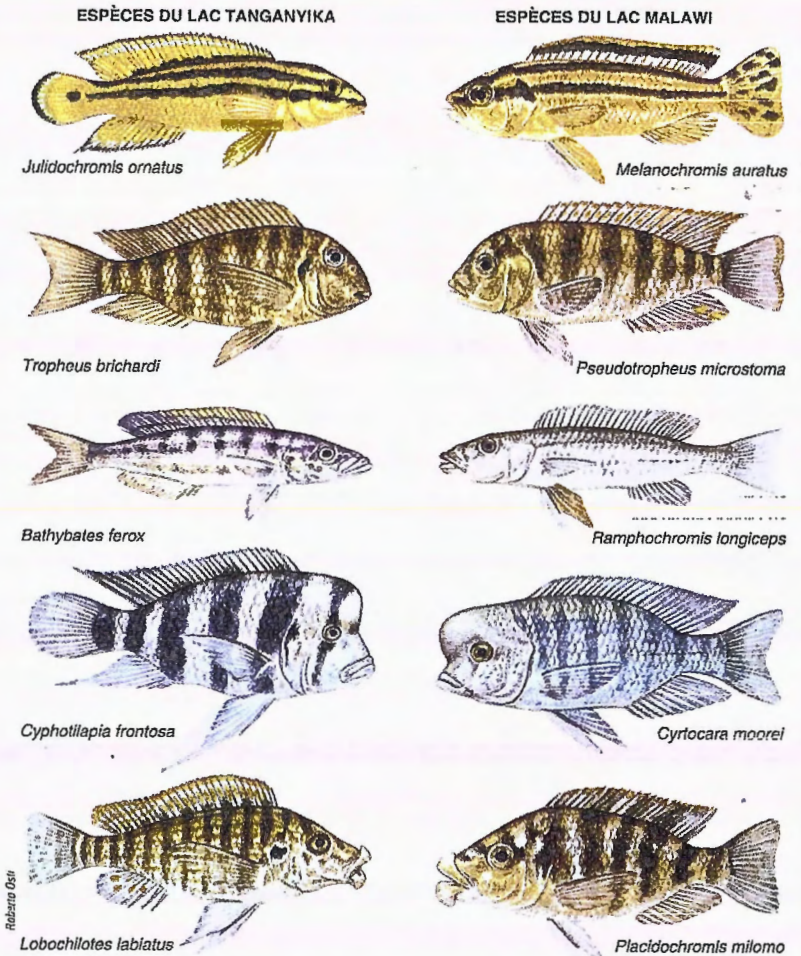
En outre, un nombre croissant d'espèces de poissons semble concerné par de tels effets, et une meilleure compréhension des mécanismes impliqués dans de tels effets environnementaux pourrait déboucher sur de nouvelles approches « propres » de contrôle du sexe, par opposition aux techniques d'inversion hormonale.

*L'étude des gènes impliqués dans ces effets environnementaux constitue déjà une première approche légère de la génomique chez le tilapia, et constitue l'une des parties majeures d'un projet européen FAIR coordonné par le CIRAD et dans lequel l'INRA intervient comme partenaire.*



# Le Tilapia, modèle complémentaire pour la Génomique Poissons

## Les Cichlidés modèles d'étude des phénomènes de spéciation



6. DES CICHLIDÉS faiblement apparentés des lacs Tanganyika et Malawi sont devenus semblables en occupant des niches écologiques similaires. Tous les cichlidés du lac Malawi sont plus étroitement apparentés entre eux qu'ils ne le sont aux cichlidés du lac Tanganyika. Les ressemblances morphologiques ne sont parfois que faiblement corrélées à la parenté génétique ou à la lignée évolutive.



# **Le Tilapia Modèle complémentaire pour la génomique Poissons**

## **2. Le tilapia et ses spécificités biologiques : appartenance à la famille des Cichlidés**

Les Tilapias appartiennent à la grande famille des Cichlidés, sur laquelle sont entrepris de nombreux travaux sur la **Biodiversité** les phénomènes de **spéciation** et d'**évolution**, en particulier dans les **grands lacs africains**.

Les outils génomiques obtenus chez le tilapia devraient être rapidement utilisables chez d'autres espèces de la même famille des Cichlidés (voir en particulier, la conservation des groupes de synténie dans la famille des Salmonidés), et donc générer de nouvelles possibilités de collaboration et de projets en Afrique.

# Le Tilapia, modèle complémentaire pour la Génomique Poissons

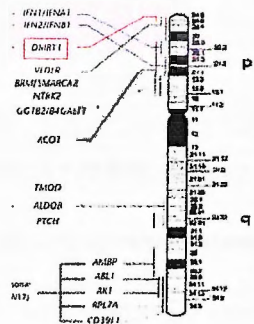
## ■ Outils spécifiques déjà acquis pour de la génomique Tilapia

Projet CEE FAIR, 2 ATP-CIRAD en collaboration INRA

AGTGAGTGAGTGAGTG

μsatellites

→ 133 ATP-HIT + banques données



Grpes de liaisons → 30 ATP-HIT + banques données

Gènes clonés =  
EST potentielles

→ 11

FAIR CEE (HDC)  
+ Dim. croissance

Banques de cDNA

→ 4

FAIR CEE (HDC)

Souches / Esp. divergeantes

→ 3 + Hybrides

ATP-HIT

Géniteurs à descendance  
monosexes

→ Mâles XX, YY, NéoFemelles ZZ



## **Le Tilapia Modèle complémentaire pour la génomique Poissons**

**Outils spécifiques déjà acquis par le CIRAD pour la génomique tilapia.**

A travers les différents projets en cours faisant appel à des techniques de biologie moléculaire (essentiellement ATP-CIRAD HIT et projet CEE-FAIR), un certain nombre d'outils techniques et/ou biologiques utilisables pour la génomique tilapia sont d'ores et déjà acquis par l'unité aquaculture du CIRAD-EMVT, en collaboration avec l'INRA.

## **PARTENAIRES**

- **Cartographie comparée: appui sur le génome humain, la Truite Arc en ciel (INRA) et autres espèces**
- **Expertise du Génoplate (INRA/CIRAD/IRD/CNRS, Privés) et appui technologique du futur Genopôle de Montpellier**
- **Université de Stirling/Swansea (Ecosse/Pays de Galles)**
- **Collaboration avec Kocher (USA) : en cours sur les microsatellites et les groupes de liaison (ATP-HIT Tilapia) chez le tilapia**
- **Centre de Biotechnologie à Chiang Mai (Thaïlande)**



## **La génomique du tilapia : intégration dans un programme génomique poissons et partenaires**

Tous les Programmes de génomique reposent sur des réseaux de compétences et des pôles techniques. Le Projet Génomique Tilapia doit nécessairement s'intégrer :

- à l'intérieur de dispositifs nationaux déjà en place (Laboratoire de Génétique des Poissons de l'INRA pour la cartographie du génome) ou démarrant (Station SCRIBE à Rennes, pour la composante poisson du programme de génomique fonctionnelle, ASTEROGAR, démarrant dès l'automne 1999).
- Et/ou à l'intérieur de pôles techniques et de compétences (Génoplateau, et futur Génopôle à Montpellier).
- Mais aussi prendre en compte les compétences européennes sur le tilapia (entre autres pour déposer des projets communs à la CE), mais aussi internationales (Etats-Unis, Israël)

Un tel programme doit aussi dès le départ prendre en compte les sollicitations extérieures (actuelles et futures), et pour cela pouvoir s'appuyer sur des structures des pays du Sud, l'idéal étant évidemment de participer à la mise en place ou au développement de certaines d'entre elles (cas de la Thaïlande où devrait se mettre en place un Centre de Biotechnologie impliquant le CIRAD).

## ANNEXES

1. Projets de Biotechnologie et/ou de Génomique sur lesquels l'Unité Aquaculture du CIRAD-EMVT est sollicité
2. Le projet ASTEROGGER de l'INRA



# Projets Biotechnologies/Génomique dont Tilapia

3

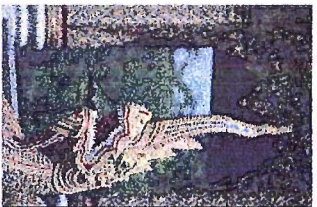


Chambre de Commerce  
Franco-Thai

Projet du

**BIOTECHNOLOGY  
PARK**

Chiang Mai,  
Thaïlande



Univ. Stellenbosch, Univ. Swansea,

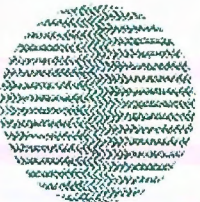
Afrique du Sud Pays de Galles



?

**Molecular genetics of Tilapia  
in Southern Africa**

**Workshop en Nov. 1999**



**INA**

1

Hunan Normal  
Univ. Chine



**Tilapia: Contrôle du Sexe,  
Gestion de souches, Sélection  
génétique, utilisation d'outils  
moléculaires**

## **Projets Biotechnologies / Génomiques sur lesquels le CIRAD-EMVT Aquaculture est sollicité**

Les biotechnologies et/ou approches liées à la génomique font d'ores et déjà l'objet de demandes de la part de nos partenaires du Sud.

En ce qui concerne le tilapia, l'unité aquaculture est actuellement sollicitée sur 3 projets « biotechnologies » par des pays du Sud :

- Participation à la mise en place d'un Centre de Biotechnologie en Thaïlande (CIRAD/. Le programme et/ou compétences affichées dans le cadre de ce projet reposent essentiellement sur les productions végétales. En ce qui concerne les productions animales, les compétences et/ou programmes proposés concernent essentiellement le tilapia. La Thaïlande est le 3<sup>ème</sup> pays producteur de tilapia dans le monde.
- Un programme franco-chinois a été mis en place en collaboration avec l'INRA, et consiste pour la première année 1999, en visites réciproques des 2 Instituts français par les partenaires chinois de l'université chinoise (Hunan Normal University). Une mission de 15 jours des 2 partenaires français Y. Guigen (INRA) et J.F. Baroiller (CIRAD) aura lieu fin Octobre 1999, et permettra de discuter de la mise en place d'un projet sur le contrôle du sexe chez l'Anguille des Rizières et le tilapia, avec utilisations d'approches moléculaires. La Chine est le premier producteur
- L'Université de Stellenbosch, en Afrique du Sud démarre un grand programme d'amélioration génétique et contrôle du sexe chez le tilapia, assisté par des marqueurs génétiques. Le CIRAD-EMVT-aquaculture a été sollicité pour faire partie de ce projet. Une mission est prévue fin 1999 pour finaliser un tel projet.



# Grand Projet Transversal "ASTEROGER"

## Analyse Systématique des Transcrits en vue de l'Etablissement d'un Répertoire Ordonné de Gènes pour l'Etude de leur Régulation

### Résumé

Dans la perspective de la description du transcriptome complet –c'est-à-dire de l'ensemble des transcrits correspondants à un génome- le projet "ASTEROGER" a pour but de concevoir, construire et exploiter un répertoire ordonné de gènes correspondant à une fraction significative (50 000 gènes) du transcriptome pour chacune des principales espèces animales d'élevage (bovins, porcins, poulet, truie).

Ce répertoire permettra de développer un ensemble d'outils moléculaires spécifiques progressivement mis à la disposition des généticiens, physiologistes, pathologistes, zootechniciens, et technologues pour leur permettre d'accéder à l'analyse simultanée du plus grand nombre de gènes mais aussi pour favoriser l'émergence de nouveaux moyens de diagnostic, contrôle, traçabilité, etc., destinés à l'ensemble des acteurs des filières des productions animales et des technologies de transformation.

Ce projet transversal suppose la mise en commun de moyens humains et logistiques conséquents qui pourraient être consentis par les différents départements concernés. Il comprend 3 étapes mettant en jeu des compétences complémentaires et permettant de déboucher sur 2 grands axes d'applications directes :

1. Construction de banques normalisées d'ADNc, multi-tissus (une par espèce),
2. Séquençage systématique de l'extrémité des clones d'ADNc pour générer des étiquettes (EST),
3. Confection et développement des outils moléculaires d'analyse transcriptionnelle (filtres haute densité, microarrays, biopuces, ...).

Les applications directes concernent :

1. La génétique (cartographie et polymorphisme des gènes exprimés, biodiversité et sélection,...),
2. La physio-pathologie, au sens large (identification de gènes ou de groupes de gènes d'intérêt par criblage différentiel, profils transcriptionnels...).

### 1. La génomique animale à l'INRA

La "génomique" se caractérise par l'acquisition systématique de données de structure et de fonction sur l'ensemble d'un génome, en utilisant des techniques à haut débit, dans le cadre de réseaux ou de consortium ; les travaux en cours sur le génome humain illustrent parfaitement ce concept. Le but de cette approche systématique est de transformer les informations de séquence en données biologiques.

Chez les animaux domestiques, le développement des cartes génétiques, réalisé dans la suite des travaux de génétique humaine, a atteint et même dépassé aujourd'hui les objectifs que l'INRA s'était assignés au lancement de ce programme (1991). Ce succès a confirmé la justesse du choix, comme pour le génome humain, d'un investissement systématique et de la mise en place de collaborations ; celles que l'INRA a soutenues se sont développées dans le cadre des programmes financés par l'Europe. La disponibilité de marqueurs génétiques assurant une couverture complète des génomes, du porc, des bovins, du poulet et, dans une moindre mesure, des salmonidés, a permis la mise en oeuvre des applications envisagées : localisation de gènes majeurs et de QTL, introgression de gènes et sélection assistée par marqueurs, contrôle et identification de parentés, analyse de la diversité génétique. En complément aux travaux de cartographie, les outils essentiels pour l'analyse des génomes des principales espèces ont été développés : méthodes de génotypage à grande échelle (Labogena), création et caractérisation de banques de grands fragments, panels d'hybrides cellulaires et d'hybrides d'irradiation. Pendant la phase d'acquisition de ces résultats, et parallèlement à la mise en place des premières

applications de la cartographie génétique, les savoir-faire fondamentaux de la biologie moléculaire et de la biologie cellulaire se sont affirmés, dans un contexte de compétition internationale.

La connaissance des gènes de l'espèce étudiée est nécessaire, malgré les progrès faits en cartographie comparative des génomes. Si la comparaison des cartes géniques et des séquences est fondamentale pour tirer parti de l'avance acquise chez les espèces modèles ou de référence (homme, souris, tétodon, voire levure) la recherche de gènes impliqués dans les fonctions importantes pour l'élevage exige de pouvoir manipuler et caractériser les gènes de ces espèces. Ces fonctions, souvent, sont difficilement transposables ou étudiables chez l'homme ou la souris (lactation, prolificité, qualité de la viande, et même résistances aux maladies où la spécificité des relations entre hôtes et pathogènes rend hasardeuse toute approche comparative) ; les gènes homologues de plusieurs espèces peuvent être de structures proches mais être soumis à des régulations très différentes (comme en témoignent par exemple les équilibres métaboliques très différents entre le porc et l'homme) ; il est aussi fréquent que des familles multigéniques soient représentées chez des espèces apparentées par des nombres différents de représentants aux fonctions diversifiées. Ainsi, les premiers succès d'identification de gènes animaux importants - correspondant à des gènes dont l'altération provoque chez l'homme ou la souris des syndromes comparables (gènes de sensibilité à l'halothane chez le porc, gène Blad et gène culard chez les bovins) - ne resteront sans doute que des exceptions. Dans ces cas l'approche comparative s'est révélée particulièrement puissante ; elle le demeure tant qu'il s'agit de raffiner les cartes génétiques dans une région, mais est limitée par la méconnaissance des gènes présents dans l'espèce étudiée tant qu'aucun gène candidat n'est proposé par les espèces modèles.

## 2. La génomique devient fonctionnelle

Des travaux visant à identifier et répertorier des gènes impliqués dans différentes fonctions ou voies métaboliques (fonctions mammaire et ovarienne, métabolisme des lipides) ont déjà été entrepris à l'INRA en s'appuyant sur des collaborations entre équipes de généticiens et de physio-pathologistes, pour s'engager résolument dans l'analyse fonctionnelle des génomes. Au delà de l'étude de fonctions spécifiques, de nouvelles problématiques de plus en plus intégratives (biologie du développement, bien-être animal) apparaissent en réponse à des objectifs de production, eux-mêmes liés à une demande sociale qui évolue. Le caractère systématique de l'analyse des gènes que nous proposons dans ce projet permettra non seulement de répondre à l'approche intégrée mais aussi d'anticiper par rapport à de nouveaux objectifs.

La caractérisation fonctionnelle des gènes transcrits dans les principales espèces d'élevage - une caractérisation systématique et non plus limitée à certains tissus ou conditions d'intérêt - apparaît donc aujourd'hui indispensable. Cette caractérisation, composante de la génomique, a pour objet d'accumuler des informations sur la structure, l'expression, la régulation, le polymorphisme, etc. pour aboutir in fine à la fonction des gènes. Parmi les 50 à 100 000 gènes que contient le génome d'une cellule d'eucaryote supérieur, seuls 10 à 20 000 sont exprimés dans un type cellulaire donné et ce répertoire varie d'un tissu à l'autre, d'un type cellulaire à un autre, et selon le stade du développement ; de plus, les niveaux d'expression varient de 1 à 10 000 copies d'ARN messager par cellule. Pour réaliser un inventaire systématique des gènes d'un génome, il est donc nécessaire de construire une collection (banque ou génothèque) d'ADNc clonés à partir de transcrits provenant de différents tissus à différents stades du développement, et "normalisés" pour assurer une représentation aussi uniforme que possible du transcriptome. Par ailleurs, compte-tenu du caractère systématique, les techniques conventionnelles d'analyse de fonction sont inopérantes et il est nécessaire de mettre en oeuvre des stratégies nouvelles. Elles sont basées sur l'hybridation de plusieurs milliers ou dizaines de milliers de gènes (représentés par leurs ADNc ou des oligonucléotides) fixés sur un support (nylon, verre ou silicium). Selon la taille du support, la densité et la nature des dépôts, on parle de filtres haute densité (ou macroarrays), de microarrays ou de biopuces (puces à ADN, DNA chips).



### 3. Les champs d'application

#### 3.1. Vers une carte des gènes

La cartographie des étiquettes décrivant les transcriptomes des principales espèces sera un moyen d'accéder à une carte des gènes des espèces - complétant les cartes de marqueurs et faisant suite à une plus grande échelle aux travaux actuels de cartographie comparée de gènes entre plusieurs espèces.

La cartographie de ces gènes sera conduite simultanément *in silico* et expérimentalement. La disponibilité attendue des séquences humaines permet de prévoir la possibilité d'un grand nombre de localisations par voie purement informatique, combinant la comparaison des séquences et l'analyse des contextes chromosomiques. Sur le plan expérimental, une approche de cartographie physique s'appuyant sur les panels d'hybrides d'irradiation sera retenue pour des travaux de cartographie locale, tandis que de nouvelles approches fondées sur l'hybridation avec des sondes spécifiques de chromosomes ou de régions chromosomiques devront être développées. Parallèlement seront menées la recherche de polymorphisme dans les séquences de gènes pour définir des marqueurs " SNP " (Single Nucleotide Polymorphism), et la mise au point de procédés adaptés au typage à grande échelle.

#### 3.2. Identification des gènes intervenant sur la variabilité des caractères

L'identification et l'analyse structurale et fonctionnelle de gènes affectant la variabilité des caractères s'inscrivent dans la continuité du développement des cartes géniques et sont à l'interface avec la génomique fonctionnelle (transcriptome et protéome). Cette identification des gènes présents dans une région du génome (clonage positionnel), qui s'appuie aujourd'hui sur les cartes existantes et la cartographie comparée, bénéficiera de la connaissance des gènes exprimés dans l'espèce et de leur localisation. En attendant une description quasiment exhaustive, la connaissance du transcriptome, combinée au développement de contigs couvrant la région étudiée, facilitera l'identification des gènes présents dans ces contigs et renforcera les différentes approches actuelles comme l'"exon-trapping", le séquençage, la transfection somatique et le RNA-display (ou les méthodes équivalentes).

#### 3.3. profils d'expression

Comme indiqué précédemment, la disponibilité d'un grand nombre de séquences et de clones d'ADNc permet une approche basée sur l'hybridation de sondes complexes préparées à partir des ARN spécifiques d'un tissu ou d'une condition physiologique particulière, avec un grand nombre de clones ordonnés sur des macro- ou des microarrays. Cette approche permet l'identification de gènes ou de groupes de gènes connus ou orphelins impliqués par exemple dans une chaîne métabolique ou sous la dépendance d'un facteur de transcription comme cela a été montré chez la levure et les fibroblastes. C'est là un outil d'analyse général, quasiment universel, et commun à plusieurs équipes ou projets.

#### 3.4. Génotypage

D'autres champs d'applications s'appuyant sur la technologie et non sur les répertoires proprement dits peuvent d'ores et déjà être anticipés. À l'image de ce qui est utilisé pour analyser le polymorphisme de gènes de grande taille (p53, BRCA1 etc...), en différents points de leur séquence (scanning), des protocoles de génotypage peuvent être développés sur des supports de type DNA chips, pour analyser les combinaisons alléliques (haplotypes) en différents points d'un même gène ou de loci multi-géniques (caséines). Des stratégies comparables sont également prévisibles pour analyser l'expression et le polymorphisme de plusieurs gènes simultanément et ainsi effectuer, sur des cibles pertinentes, des études de biodiversité à grande échelle.

#### 4. Le projet : de la réalisation à l'exploitation des répertoires

##### 4.1 : étape 1, construction de banques d'ADNc multi-tissus (1999 - 2000)

Construire quatre répertoires ordonnés (un par espèce) de gènes transcrits suppose de collecter un grand nombre de séquences de type EST (50 000 au minimum, pour chaque espèce, soit 200 000 séquences) La stratégie envisagée est de construire des banques (généthèques) d'ADN complémentaires à partir d'ARNm extraits de différents tissus; ces banques multi-tissus seront "normalisées" afin d'égaliser la représentation de chaque transcrit, en limitant la redondance et donc le séquençage répété de plusieurs clones représentatifs d'un même transcrit. En complément des banques multi-tissus, des banques spécialisées seront construites, normalisées ou non, pour répondre aux besoins spécifiques des équipes concernées (e.g. : embryon, glande mammaire, cerveau, ...) et pour compléter le répertoire partiel établi à partir des banques multi-tissus. Toutes ces banques seront organisées en plaques de microtitration (96 ou 384 puits).

Cette opération sera réalisée dans les meilleurs délais en concertation entre les équipes concernées et son financement pourrait être assuré sur des crédits "Génomes et Fonctions".

##### 4.2 : étape 2, séquençage systématique (démarrage en 2000)

Une première tranche de 100 000 (4 x 25 000) séquences pourrait être réalisée par le Centre National de Séquençage dans le cadre du prochain appel d'offres; une seconde tranche d'ampleur équivalente devra être lancée dès 2002. En dehors du CNS, des collaborations sont et seront recherchées dans le cadre des appels d'offre du 5<sup>ème</sup> programme cadre européen. On peut estimer le coût global de cette phase du projet à environ 40 M€ pour les 4 x 50 000 séquences à réaliser. Le traitement des séquences obtenues ("nettoyage", contrôle et annotation des séquences) et la gestion des informations nécessiteront des moyens informatiques correctement dimensionnés et du personnel qualifié.

Dès ce stade, une première forme d'exploitation des données sera engagée, indépendamment des étapes ultérieures : elle aura pour objectif la localisation des gènes correspondants sur les chromosomes en utilisant les panels d'hybrides.

##### 4.3 : étape 3, Filtres Haute Densité, microarrays et puces

Le séquençage systématique des clones ne constitue pas un préalable au regard de certaines applications; la réalisation des filtres haute densité, en particulier, peut être menée simultanément.

L'analyse au moyen des macroarrays et des microarrays requiert des équipements spécifiques dont l'acquisition doit être envisagée : robot, imageur, stations de travail. Au delà de l'optimisation et de la standardisation des conditions d'hybridation, des collaborations seront nécessaires pour participer au développement de logiciels d'acquisition et de traitement des données et d'exploitation des résultats par des outils informatiques. Ces développements informatiques sont essentiels pour permettre au projet d'atteindre tous ses objectifs. Nous nous appuierons pour cela sur l'expérience d'équipes ayant déjà consenti un important investissement en la matière (Charles Auffray, CNRS ; Bertrand Jordan, INSERM ; Daniel Caput, SANOFI).

#### 5. Conclusions

Ce projet est ambitieux puisqu'il s'agit de collecter, séquencer et ordonner sur des supports quelques 50 000 ADN complémentaires pour chacune des 4 espèces; il permettra la création de nouveaux outils et l'acquisition de savoir-faire dont les possibilités ne sont pas épuisées par les quelques utilisations que nous avons évoquées.

L'utilisation à grande échelle de ces outils de génomique fonctionnelle n'interviendra que progressivement après le démarrage du projet. Cependant, les travaux ont commencé compte-tenu de la disponibilité du matériel biologique et de l'accès à des outils d'analyse de l'expression : programme



ROGER pour l'espèce bovine (production des fragments d'ADNc, construction de filtres et hybridation avec des sondes complexes); collection d'ADNc d'origine ovarienne porcine à Toulouse.

La construction et la mise en place de ces outils nécessiteront une collaboration étroite entre les équipes de génétique, de physio-pathologie et de zootechnie. Leur exploitation sera très dépendante des questions biologiques étudiées mais, surtout, elle permettra de les aborder sous un angle nouveau comme, par exemple, l'analyse sur un répertoire restreint (quelques dizaines) de gènes cibles connus (e.g. voies métaboliques) ou au contraire, la recherche de gènes inconnus sur l'ensemble du répertoire.